

# The AmeriFlux Cyberinfrastructure Prototype – Introduction and Applications



Gretchen Miller  
University of California - Berkeley  
Berkeley Water Center (BWC)

AmeriFlux Science Meeting  
October 17, 2006



# Outline

- Introduction to the prototype and the BWC Cyberinfrastructure Project
- Example analyses using datacube
- Comments on initial user experience
- Future improvements to the datacube
- Live example?
- Next presentation: how to use portal

# Introduction

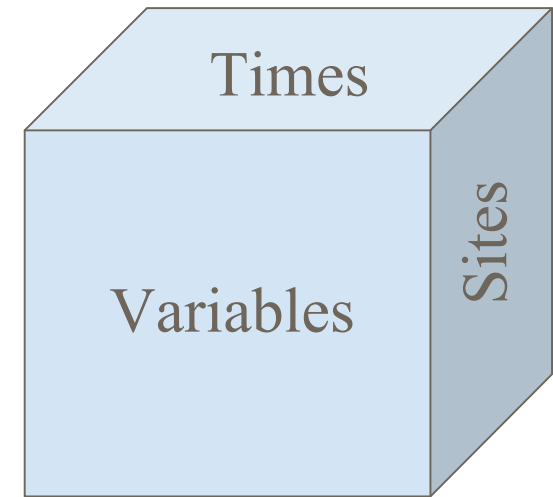
- Prototype allows easy, instant access to entire AmeriFlux database
- Designed to aid in multi-site synthesis studies
- Resulted from collaboration between the Berkeley Water Center and Microsoft



Graphics courtesy of: AmeriFlux  
and Berkeley Water Center

# What is a DataCube?


- Direct, interactive AmeriFlux database access system using Excel pivot tables
- Contains:
  - Multiple data types from
  - Multiple sites at
  - Multiple time-steps
- Automatically downloads the data you need – use now or save for later
- Connect through internet – more later...





■ Download, copy, paste, rinse and repeat

CO2 concentration	Metolius	CO2 concentration	Tonzi Ra
395			
396			39
397			39
399			39
399			39
399			39
394			39
389			39
384			3
381			38
380			38
378			38
377			37
377			37
376			37
			37
376			
376			37
379			3
383			37
386			38
389			38
392			38
393			39



# The New Way – Cubic!

■ Connect, select, and done

**Create New Data Source**

What name do you want to give your data source?

1. mycube

Select an OLAP provider for the database you want to access:

2. Microsoft OLE DB Provider for Analysis Services 9.0

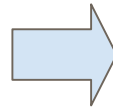
Click Connect and enter any information requested by the provider:

3.

Select the Cube that contains the data you want:

4.

☐ Save my user ID and password in the data source definition



**PivotTable and PivotChart Wizard - Layout**

Construct your PivotTable report by dragging the field buttons on the right to the diagram on the left.

Diagram:

PAGE: [ ]

Site, Datumtyp

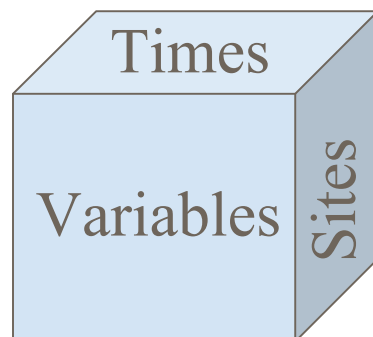
Year

Average

ROW DATA

Field List:

- Datumtyp
- Day of Year
- Value
- Exdatumtyp
- Hour of Day
- Count
- Offset
- Month of Year
- Minimum Value
- Latitude
- Year
- Maximum Value
- Longitude
- Year To Date
- Average
- Site
- Year To Month
- HasData



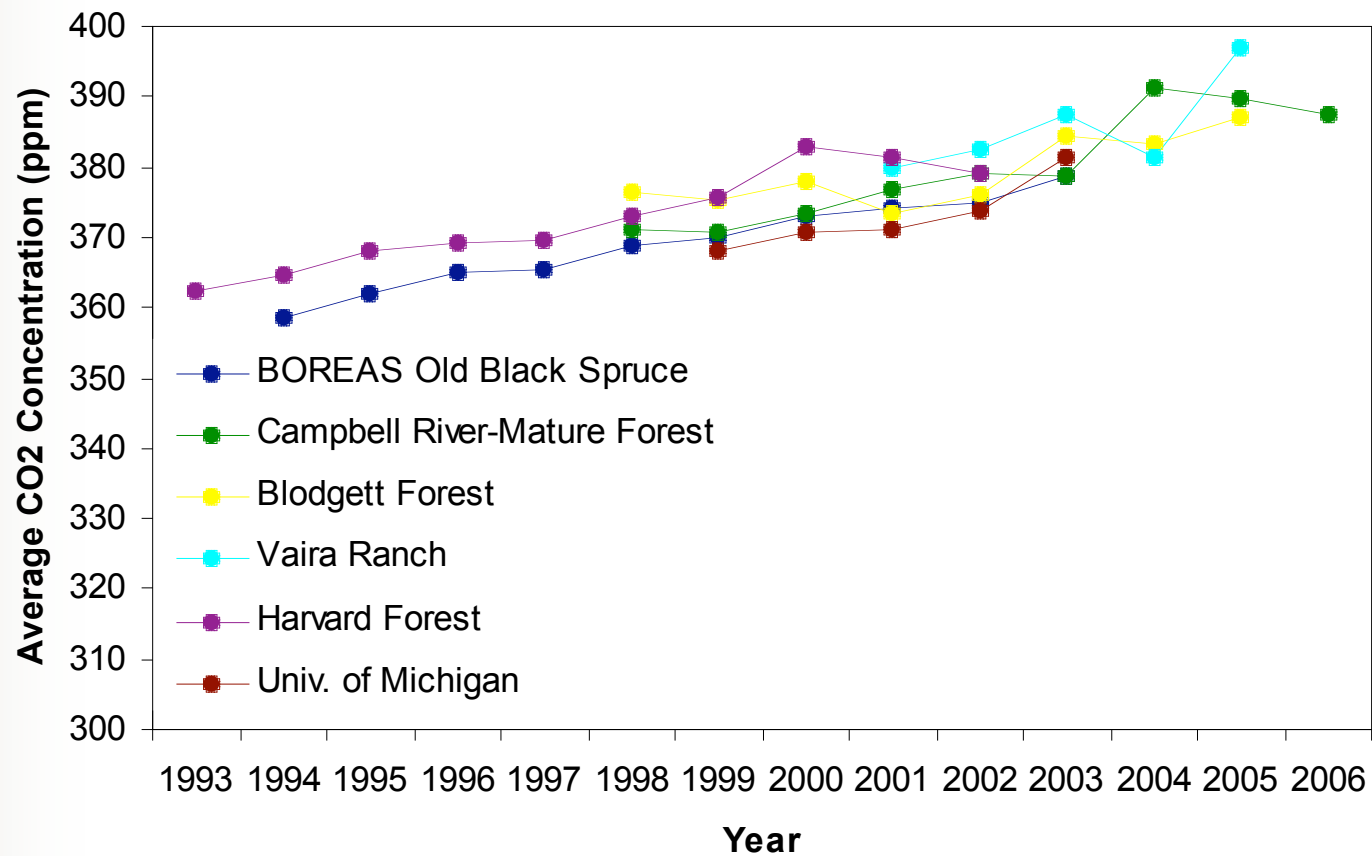
	Site				
	BOREAS NSA	Campbell River-Mature Forest	Harvard Forest	Vaira Ranch	Harvard Forest
	Average	Average		Average	Average
Year	CO2 concentration	CO2 concentration	CO2 concentration	CO2 concentration	CO2 concentration
1993					362.4899623
1994	358.4500344				364.7439787
1995	361.7871797				368.030905
1996	364.8862779				369.2956202
1997	365.2289554				369.5749438
1998	368.7482114	371.2502774	376.5199233		373.1117676
1999	369.8224038	370.5811908	375.22674		375.7325084
2000	373.1352253	373.369376	377.9258365		382.8674484
2001	374.159215	376.9836859	373.486263	379.6723482	381.5096281
2002	374.9890386	379.1057499	375.8655493	382.6557193	379.1480462
2003	378.5277665	378.6299502	384.4003365	387.5	
2004		391.2762433	383.4555272	381.5077891	
2005		389.8631304	387.2000838	397.1258357	
2006		387.5196749			
Grand Total	369.5915581	382.6706499	379.19824	385.812352	373.2583991



# Example Analyses

- Example plots shown in this presentation:
  - No need for creation of advanced spreadsheets or programs
  - Most plots took **under 15 minutes** each to create, start to finish
- Network wide examples
- Multi-site examples






# Multi-site Plots for CO<sub>2</sub>





# Steps to Create Average CO<sub>2</sub> Plot





## Before

-  Download data for each site from database
-  Cut and paste into on spreadsheet
-  Create formula to calculate yearly average for each
-  Plot
-  Repeat process for additional sites

Total Time > 2 hour, dependent on number of sites analyzed

Time to add a site ~ 30 min

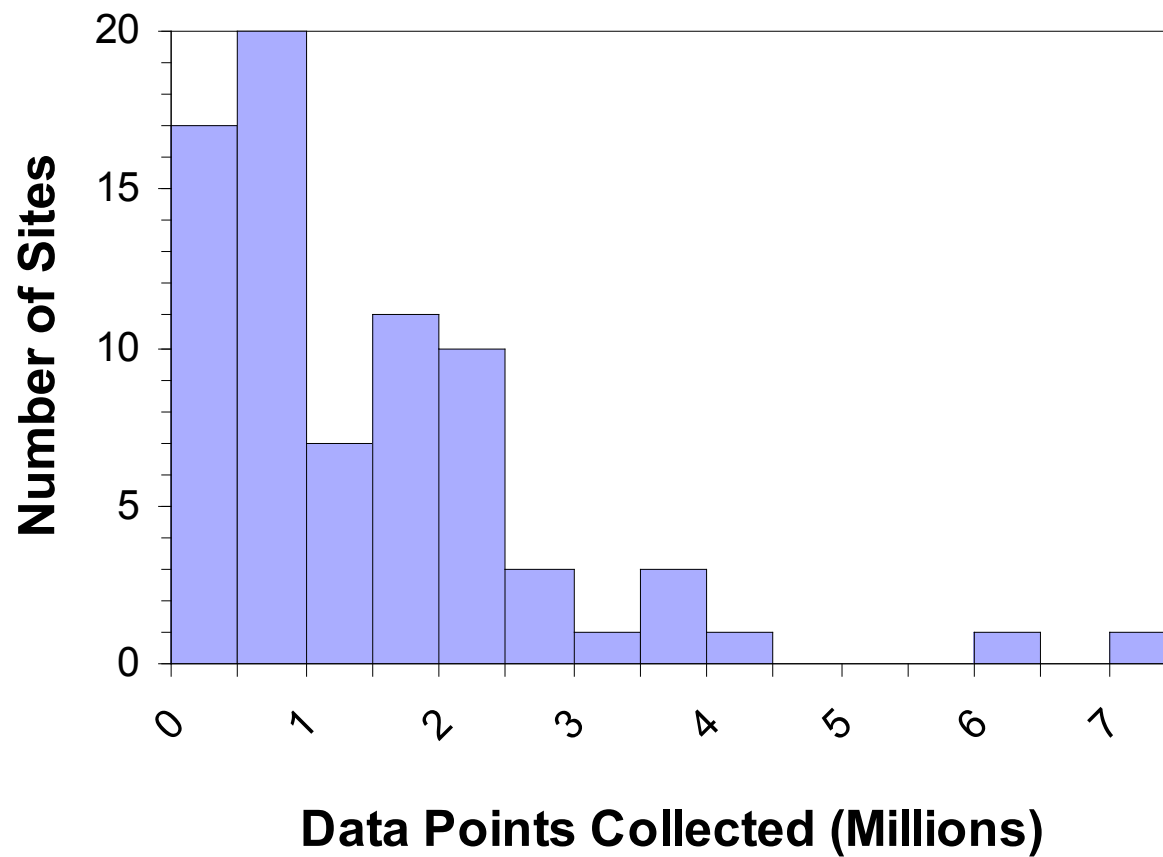
## Now

-  Connect to BWC server through Excel
-  Select desired sites, years, and “average” field
-  Display data and plot
-  Change sites by clicking twice

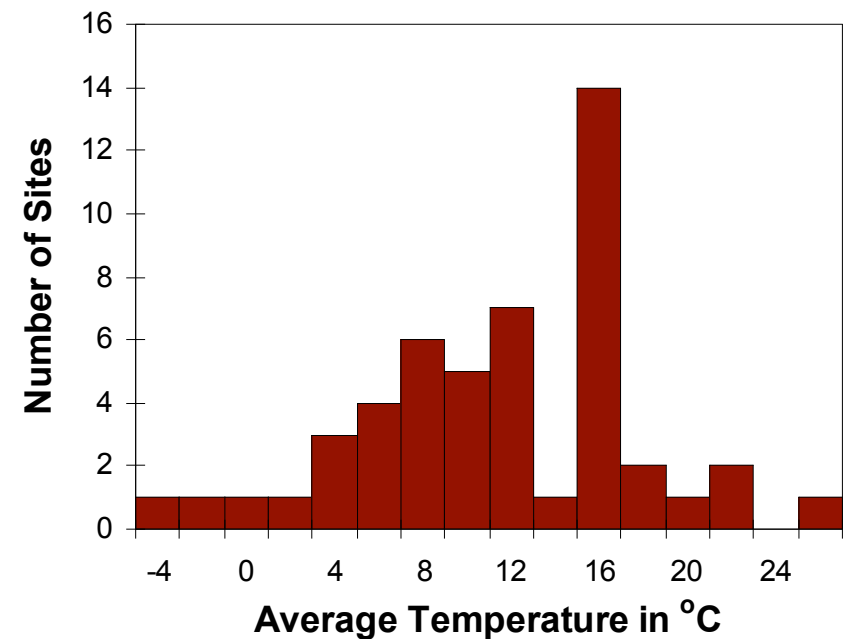
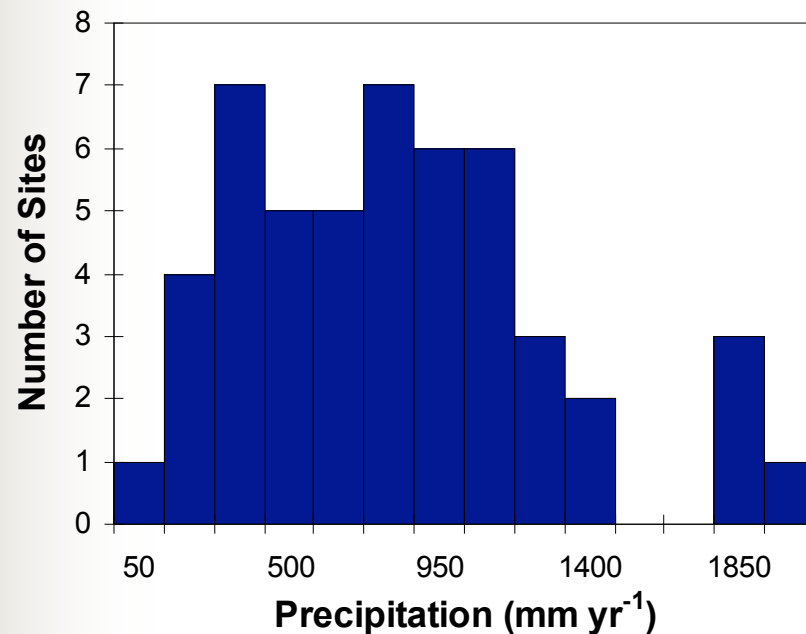
Total Time <10 min, independent of number of sites analyzed

Time to add a site < 60 sec

# System Wide Data Availability

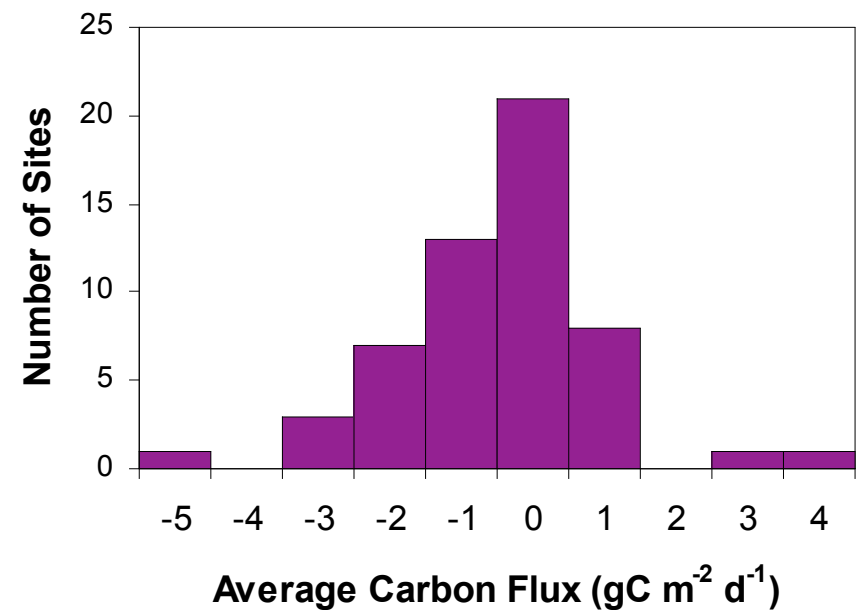
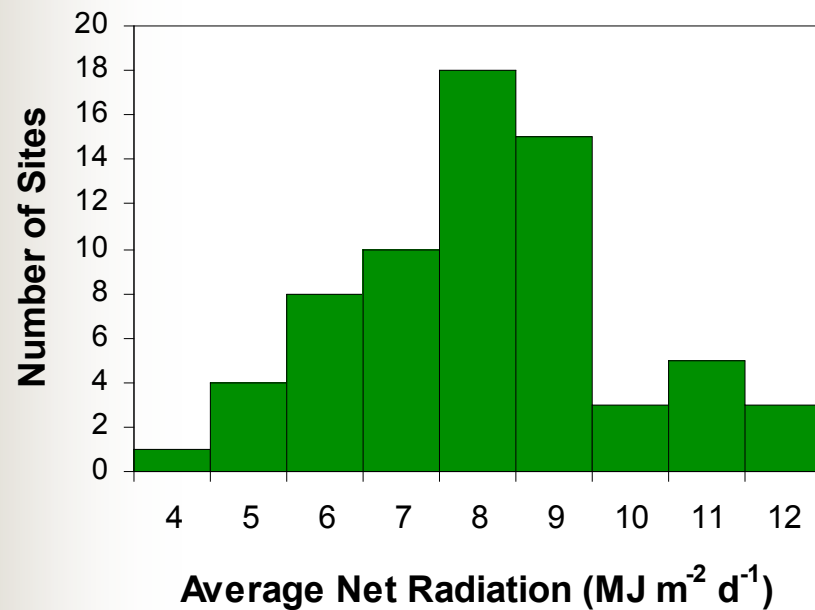


# System Wide Plots for 2004 Data



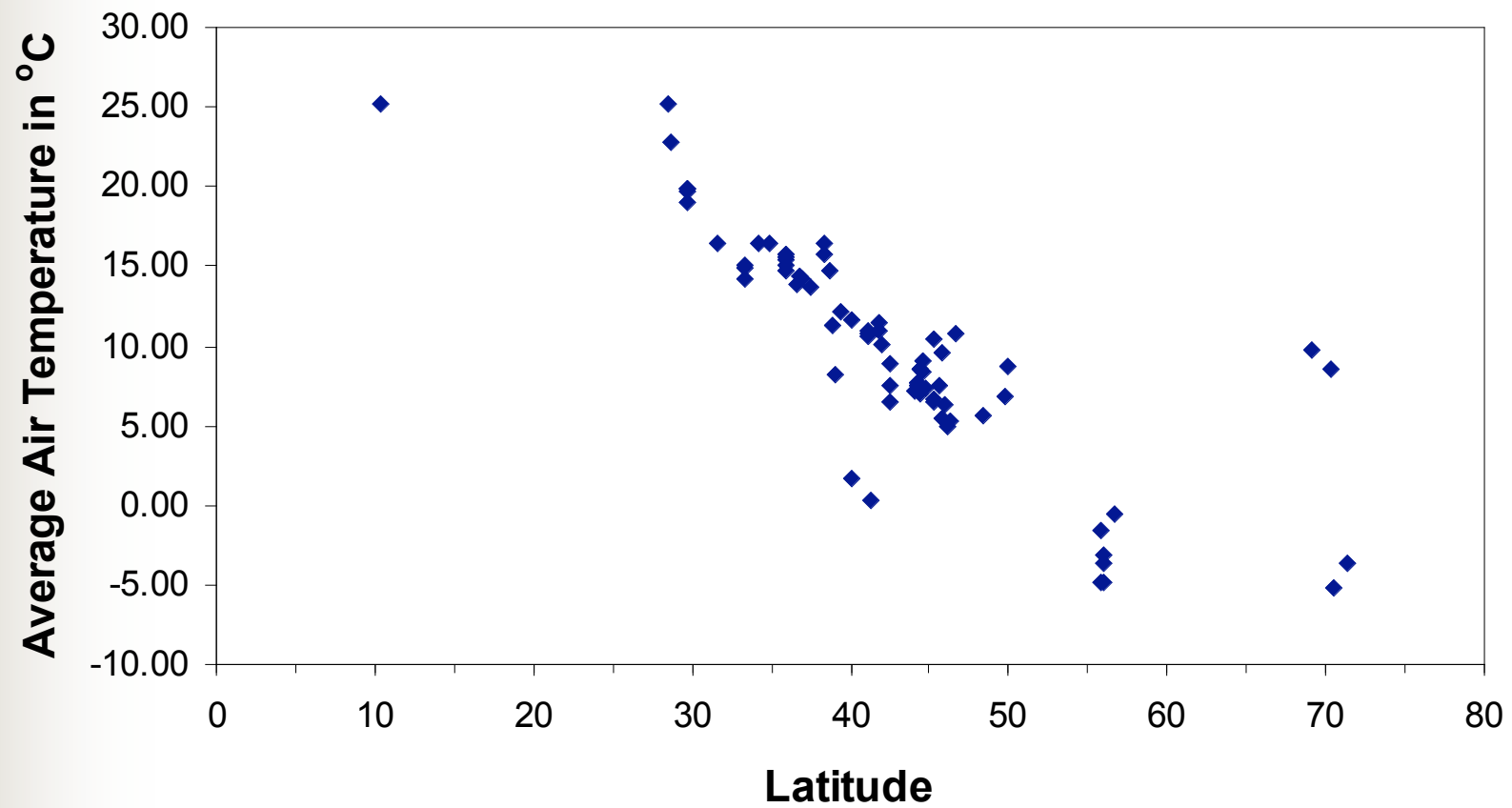
- Number of data points involved in each graph: 76 sites \* 17,520 measurements

# System Wide Plots for 2004 Data

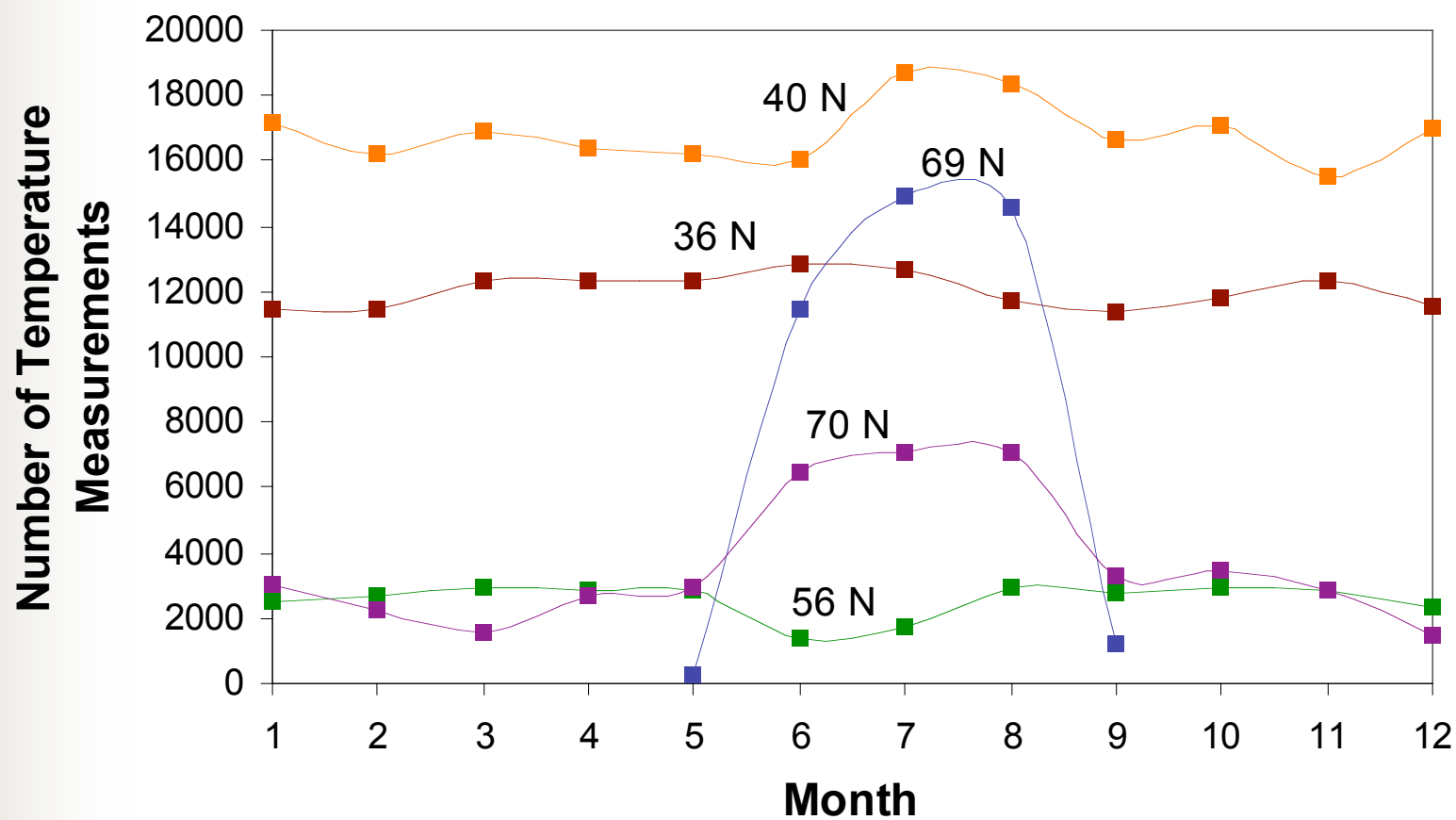




# Temperature vs. Latitude

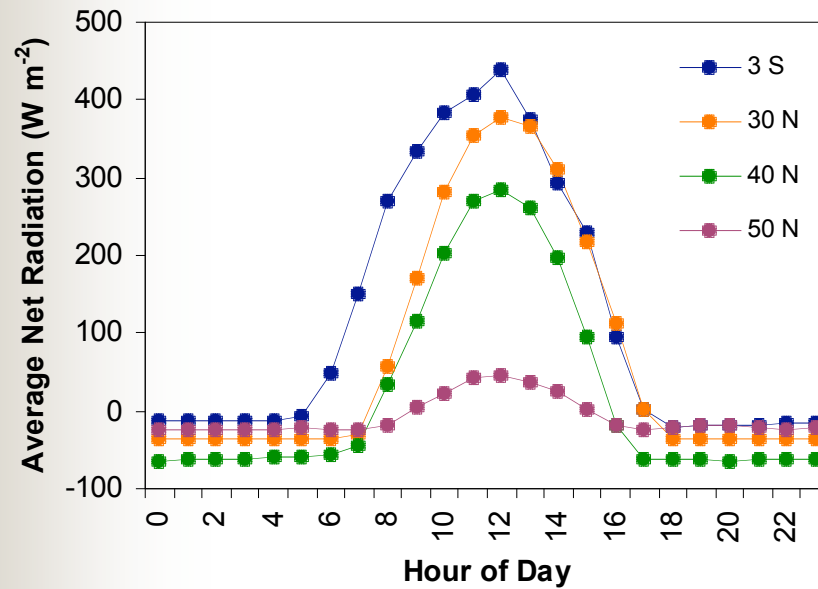


# Outliers?

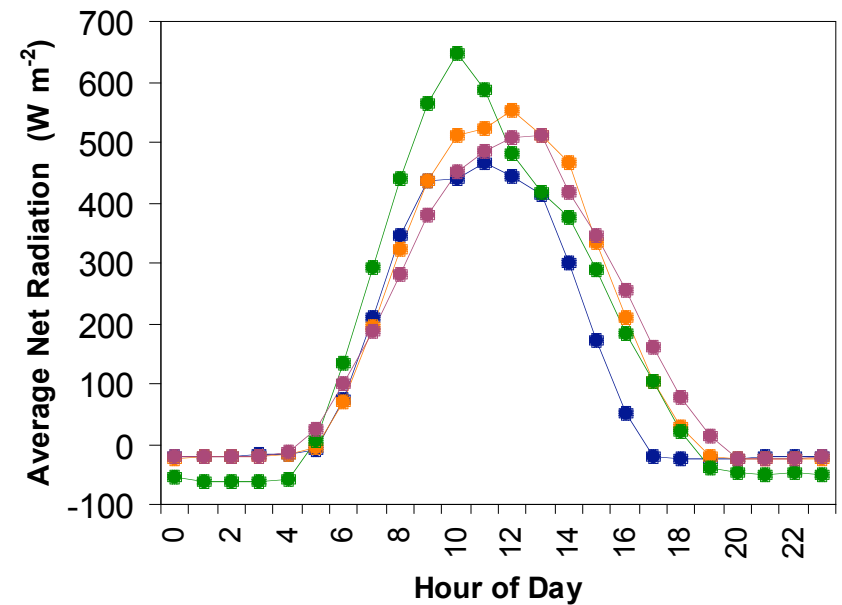


# Net Radiation and Latitude

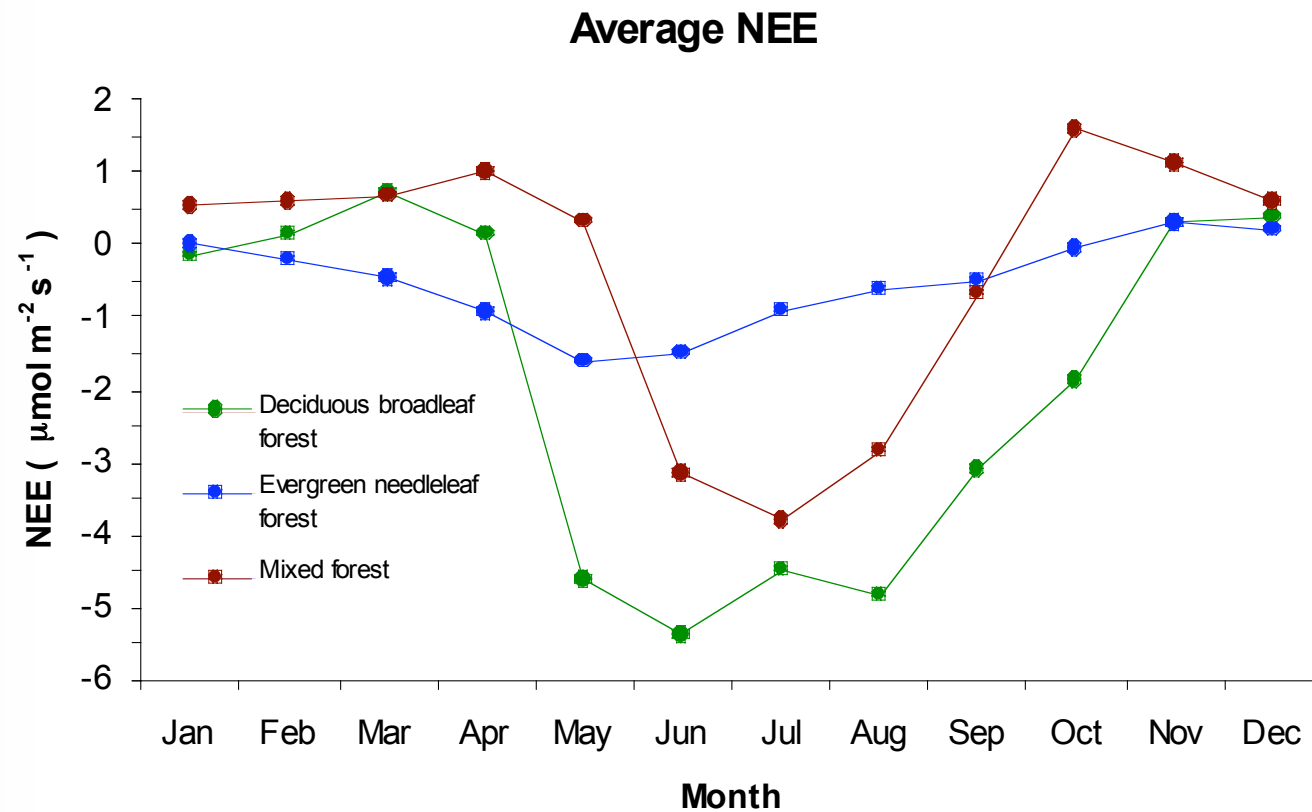
January 2004



June 2004

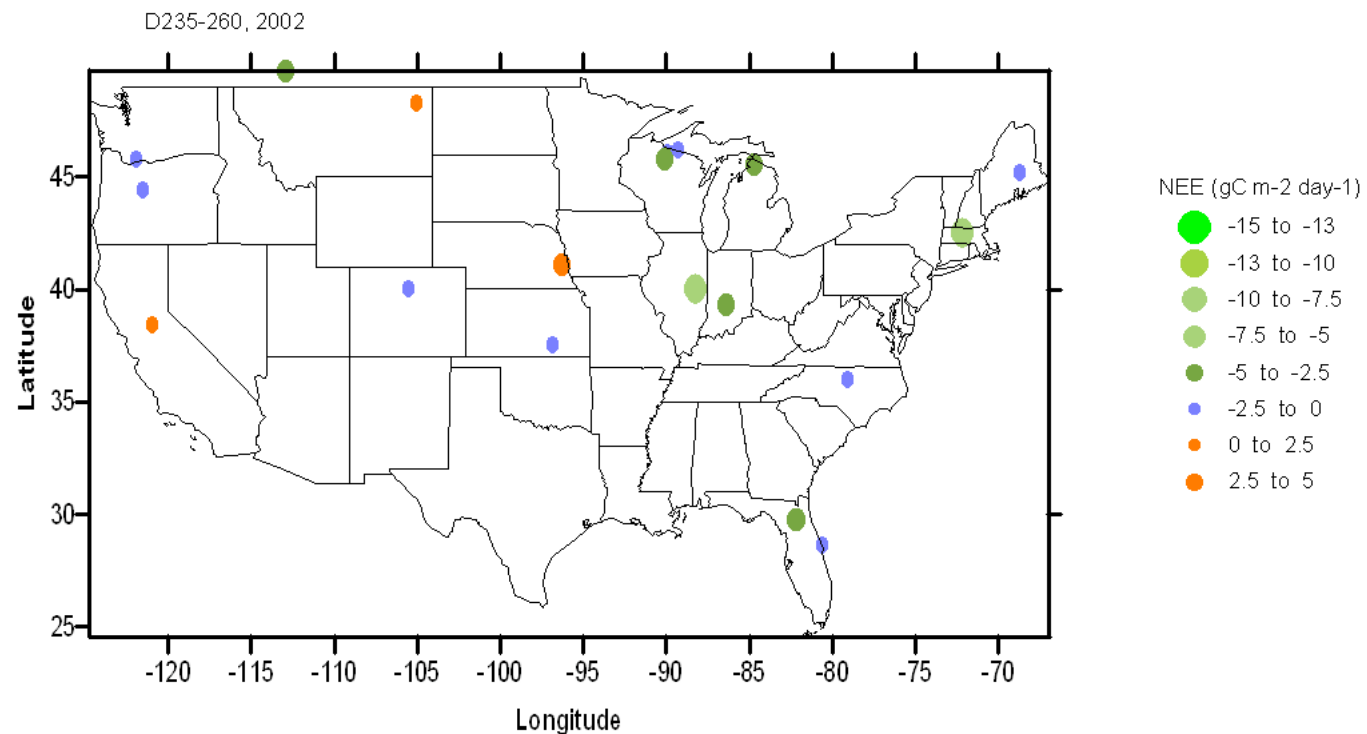


# NEE by ecosystem type





# Network NEE

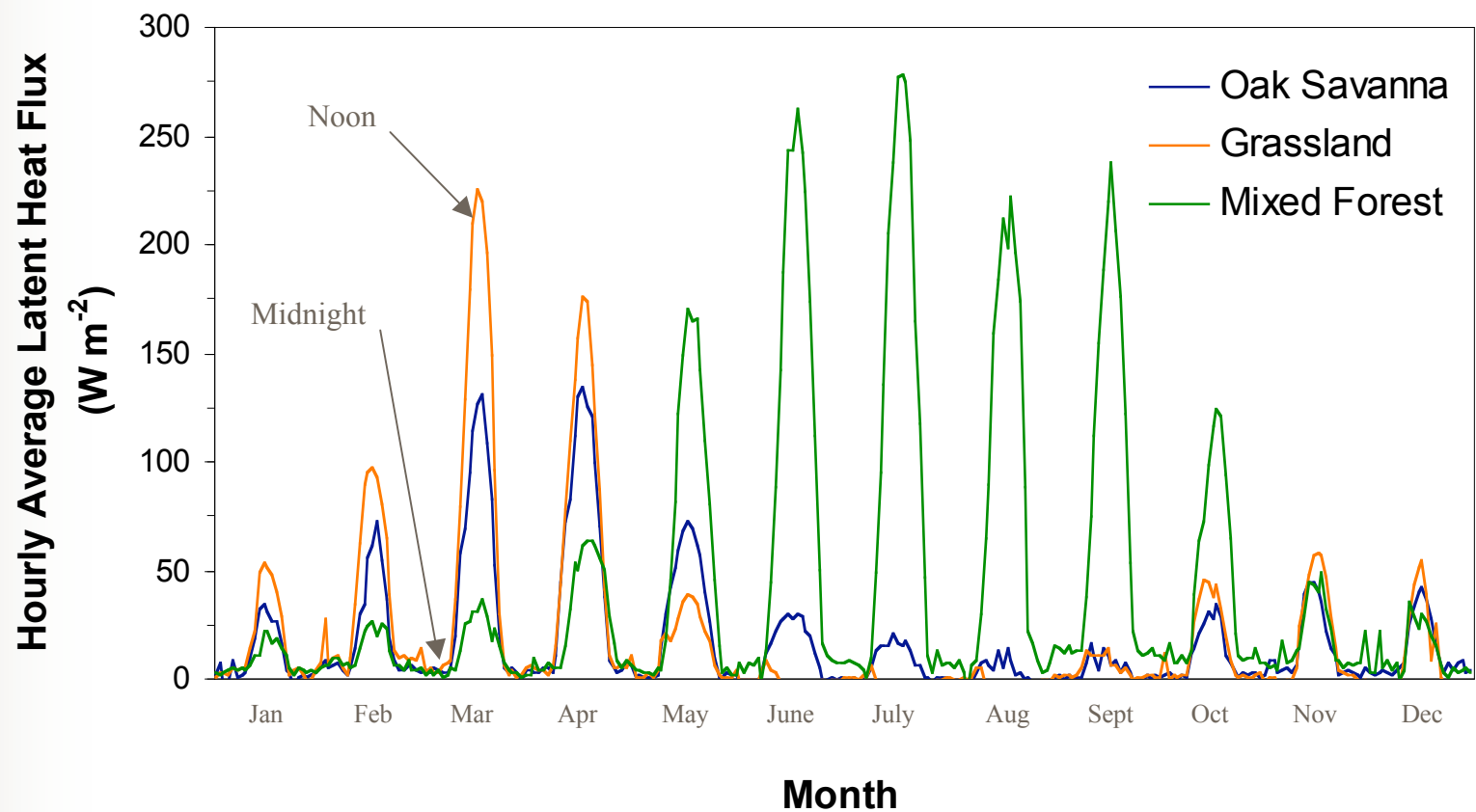


- Compiling data for figure originally took >20 man-hours
- With new method, <1 hour

Figure courtesy of D.D. Baldocchi

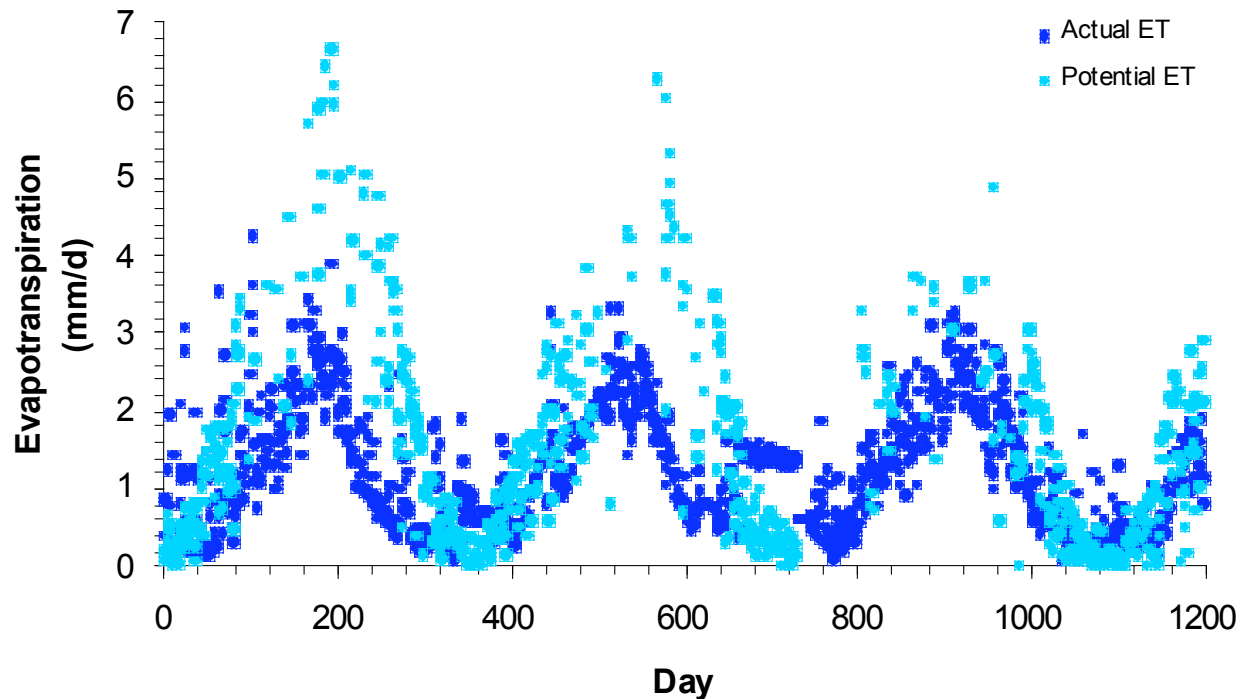
# Diurnal averages

Diurnal Pattern of LE by Month for 2004



# Using Calculated ET Values

Evapotranspiration at Metolius, OR Site



- Potential ET calculated using Priestly-Taylor equation incorporated into data cube



# Comments on initial experience

## ■ Advantages:

- Greatly simplifies data management
- Decreases time spent performing repetitive calculations
- Easier to view data from multiple sites
- Quickly switches between “levels” of data
  - Change between yearly, monthly, daily, and hourly sums or averages in seconds
- Less error prone
  - More eyes on the data
  - Fewer manual operations and/or programming





# Comments on initial experience

## ■ Challenges:

- Many analyses need gap filled data
- Some ways of displaying data can mask missing or unreasonable values

## ■ Conclusions:

- Think MORE, Work LESS
- Give poor graduate students more time to do their field work



# Future Improvements

- Gap filled data incorporated directly into portal
  - Currently working with a team from FLUXNET to incorporate several methods into cubes
- Built-in unit conversions and common calculations – suggestions?
- Better meta-data and ecological data handling
  - Ecosystem type, soil texture, lat/long, instrument type
  - LAI, spectral reflectance,  $V_{\text{cmax}}$ , predawn water potential, soil respiration
- Alternative plotting and statistics software



# Acknowledgements

- Microsoft Berkeley Water Center Project
  - Catherine van Ingen
- BWC AmeriFlux Working Group
  - Deb Agarwal, Dennis Baldocchi, Bev Law, Matt Rodriguez, Monte Goode
- Berkeley Water Center
  - Yoram Rubin, Susan Hubbard
- CarboEurope Data Team
  - Dario Papale, Markus Reichstein
- NSF Graduate Student Fellowship Program



# Live Example